Time and Frequency Domain Optimization with Shift, Convolution and Smoothness in Factor Analysis Type Decompositions

Madsen, Kristoffer Hougaard; Hansen, Lars Kai; Mørup, Morten

Publication date:
2009

Document Version
Early version, also known as pre-print

TIME AND FREQUENCY DOMAIN OPTIMIZATION WITH SHIFT, CONVOLUTION AND SMOOTHNESS IN FACTOR ANALYSIS TYPE DECOMPOSITIONS.

Kristoffer H. Madsen and Lars K. Hansen and Morten Mørup

Technical University of Denmark
Informatics and Mathematical Modelling
Richard Petersens Plads, Building 321
DK-2800 Kgs. Lyngby, Denmark
email: {khm,lkh,mm}@imm.dtu.dk

ABSTRACT

We propose the Time Frequency Gradient Method (TFGM) which forms a framework for optimization of models that are constrained in the time domain while having efficient representations in the frequency domain. Since the constraints in the time domain in general are not transparent in a frequency representation we demonstrate how the class of objective functions that are separable in either time or frequency instances allow the gradient in the time or frequency domain to be converted to the opposing domain. We further demonstrate the usefulness of this framework for three different models; Shifted Non-negative Matrix Factorization, Convolutive Sparse Coding as well as Smooth and Sparse Matrix Factorization. Matlab implementation of the proposed algorithms are available for download at
`www.erpwavelab.org`.

1. INTRODUCTION

Factor analysis [1] is widely used to reconstruct multiple latent effects from mixtures based on the model

$$x_{m,n} \approx \sum_d a_{m,d} s_{d,n}. \tag{1}$$

However, the decomposition is not unique since $\widehat{\mathbf{A}} = \mathbf{AQ}$ and $\widehat{\mathbf{S}} = \mathbf{Q}^{-1}\mathbf{S}$ yield the same approximation as $\mathbf{A}, \mathbf{S}$. Consequently, constraints have been imposed such as Varimax rotation for Principal Component Analysis (PCA) [2], statistical independence of the sources $\mathbf{S}$ as in Independent Component Analysis (ICA) [3, 4]. A related strategy is sparse coding where the objective of minimizing the error is combined with a term enforcing sparsity of $\mathbf{S}$ [5].

Factor analysis in the setting of ICA is often illustrated by the so-called cocktail party problem. Here mixtures of several speakers are recorded in a number of microphones forming the measured signal $\mathbf{X}$. The task is to recover the original speech signals $\mathbf{S}$. Harshman and Hong [6] proposed a generalization of the factor model in which the source signals may be delayed on arrival at the the sensors. The model is called shifted factor analysis (SFA), and reads

$$x_{m,n} \approx \sum_d a_{m,d} s_{d,n-\tau_{m,d}}. \tag{2}$$

equivalent models have been proposed in [4, 7, 8]. In real acoustic environments we expect echoes due to multiple signal propagation paths, that arise from reflections off surfaces. To account for general delayed mixing, factor analysis models have been further generalized to convolutive mixtures, see e.g., [9, 10, 11, 12, 13]

$$x_{m,n} \approx \sum_{\tau,d} a^{\tau}_{m,d} s_{d,n-\tau}. \tag{3}$$

Here $\mathbf{A}^{\tau}$ is a filter that accounts for the presence of each source in the sensors at time delay τ. The shifted factor model, thus is a special case of the convolutive model where the filter coefficients $a^{\tau}_{m,d} = a_{m,d}$ if $\tau_{m,d} = \tau$ otherwise $a^{\tau}_{m,d} = 0$. A very comprehensive survey of methods for convolutive ICA is found in [14].

It is well known that the frequency representation of time series data provides an efficient framework to account for models with shifts, convolutions and smoothness due to the following basic observations

- Shift and convolution in the time domain corresponds to multiplication in the frequency domain.

- Smoothness of signals in the time domain correspond to low frequency representations in the frequency domain.

The presence of efficient FFT algorithms for transformation between the time and frequency domain has a limited computational cost ($\mathcal{O}(n \log n)$). Unfortunately, models are often constrained in the time domain in a way that is not transparent in a frequency representation and this may hamper optimization in the frequency domain. To improve component identifiability we impose constraints such sparseness, non-negativity, or smoothness. While shifts and convolution are efficiently implemented in the frequency domain constraints in the form of non-negativity, sparseness and smoothness are typically defined in the time domain without an explicit representation in the frequency domain.

We will here propose a gradient based technique that facilitates taking advantage of a frequency representation while accounting for constraints in the time-domain. We will demonstrate the usefulness of this approach on the following models

- **Shifted Non-negative Matrix Factorization**, the data and model parameters are constrained non-negative in the time domain while temporal shifts are represented efficiently through the frequency domain representation.

- **Convolutive Sparse Coding**, here convolution is efficiently implemented through multiplication in the frequency domain while the filter length in the time domain constrains regions of $\mathbf{A}^\tau$ to zero. Furthermore, the sparseness imposed on $\mathbf{S}$ resides in the time domain.

- **Sparse and Smooth Matrix Factorization**, while smoothness constraints can efficiently be implemented in the frequency domain, sparseness constraint resides in the time domain.

The paper is structured as follows. We will first describe the proposed Time Frequency Gradient Method (TFGM) that combines time and frequency domain optimization. Based on this approach we derive efficient algorithms for the three models above and demonstrate their use on simulated as well as real data. We have previously used the TFGM for Shifted Non-negative Matrix in [17], here the aim is to formalize the approach as well as generalize to broader classes of models.

2. NOTATION

In the following $\mathbf{u}$ and $\mathbf{U}$ will denote a vector and matrix in the time domain, while $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{U}}$ denotes the corresponding vector and matrix in the frequency domain. Furthermore, $\widetilde{\mathbf{U}}^H$ denotes the conjugate transpose of $\widetilde{\mathbf{U}}$ while $\mathbf{U} \bullet \mathbf{V}$ denotes the direct product, i.e. element-wise multiplication. $\widetilde{\mathbf{U}}^{(f)} = \mathbf{U} \bullet e^{-i2\pi \frac{f-1}{N} \boldsymbol{\tau}}$ where $e^{-i2\pi \frac{f-1}{N} \boldsymbol{\tau}}$ denotes element wise raising the elements, i.e.$(e^{-i2\pi \frac{f-1}{N} \boldsymbol{\tau}})_{n,d} = e^{-i2\pi \frac{f-1}{N} \tau_{n,d}}$. Also, let $\mathbf{u}_d$ denote the d^{th} column and $u_{n,d}$ a given element of $\mathbf{U}$. Finally, let $\|\mathbf{U}\|_F = \sqrt{\sum_{m,n} |u_{m,n}|^2}$ be the Frobenius-norm and $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the discrete Fourier and inverse Fourier transform, i.e.

$$\mathcal{F}(s): \quad \widetilde{s}_k = \sum_{n=0}^{N-1} s_n e^{-i\frac{2\pi k}{N} n}$$

$$\mathcal{F}^{-1}(\widetilde{s}): \quad s_n = \frac{1}{N} \sum_{k=0}^{N-1} \widetilde{s}_k e^{i\frac{2\pi k}{N} n}.$$

3. METHODS AND RESULTS

We will presently consider objective functions, C, of the form

$$C = \sum_t f_t(x_t) + \frac{1}{N} \sum_f g_f(\widetilde{x}_f), \tag{4}$$

where f_t and g_f are real valued functions of the real and complex variables x_t and $\widetilde{x}_f$ such that $\widetilde{\mathbf{x}} = \mathcal{F}(\mathbf{x})$. Thus, we require the objective functions to be separable in either the time or frequency domain. The gradient with respect to x_t and $\widetilde{x}_f$ of objective functions satisfying (4) can be written as

$$\begin{aligned}
\frac{\partial C}{\partial x_t} &= f_t'(x_t) + \frac{1}{N} \sum_f g_f'(\widetilde{x}_f) e^{i2\pi f t} \\
&= f_t'(x_t) + \mathcal{F}^{-1}(\mathbf{g}')_t \\
\frac{\partial C}{\partial \widetilde{x}_f} &= \sum_t f'(x_t) e^{-i2\pi f t} + g_f'(\widetilde{x}_f) \\
&= \mathcal{F}(\mathbf{f}')_f + g_f'(\widetilde{x}_f)
\end{aligned}$$

Thus, the gradient of the objectives can be converted arbitrarily between the time and frequency domain enabling what we will denote the Time-Frequency Gradient Method (TFGM). The crux of this property follows from the separability into sums over time or frequency instances (t or f). We note, that from Parseval's identity we can arbitrarily state the least square

objective in a form satisfying (4) both in the time and frequency domain, i.e.,

$$\sum_n \|\mathbf{x}_n\|_F^2 = \frac{1}{N}\sum_f \|\mathbf{x}_f\|_F^2. \quad (5)$$

We note that a variable *which***x** is updated in the frequency domain has to remain real when applying the inverse DFT. For $\mathcal{F}^{-1}(\mathbf{g})$ to be real valued the following has to hold

$$\mathbf{g}_{N-f+1} = \mathbf{g}_f^*,$$

where * denotes the complex conjugate. This constraint is enforced by only considering the first $\lfloor N/2 \rfloor + 1$ frequencies, i.e. up to the Nyquist frequency, while setting the functions of the remaining frequencies according to (**??**).

3.1. Shifted Non-negative Matrix Factorization

A popular approach for enforcing non-negativity is the use of multiplicative updates as introduced in [15, 16]. Given a cost function $C(\mathbf{S})$ over the non-negative variables $\mathbf{S}$, define $\frac{\partial C(\mathbf{S})^+}{\partial s_{d,n}}$ and $\frac{\partial C(\mathbf{S})^-}{\partial s_{d,n}}$ as the positive and negative part of the derivative with respect to $\mathbf{S}_{d,m}$. Then the multiplicative update has the following form

$$s_{d,n} \leftarrow s_{d,n}\left(\frac{\frac{\partial C(\mathbf{S})^-}{\partial s_{d,n}}}{\frac{\partial C(\mathbf{S})^+}{\partial s_{d,n}}}\right)^{\alpha}.$$

A small constant $\varepsilon = 10^{-9}$ is added to the numerator and denominator to avoid division by zero or forcing $s_{d,n}$ to zero. If the gradient is positive $\frac{\partial C(\mathbf{S})^+}{\partial s_{d,n}} > \frac{\partial C(\mathbf{S})^-}{\partial s_{d,n}}$, hence, $s_{d,n}$ will decrease and vice versa if the gradient is negative. Thus, there is a one-to-one relation between fixed points and points where the gradient is zero. The attractive property of multiplicative updates is that they automatically ensure non-negativity as the updates are based on multiplication, division, and raising to the power of purely non-negative variables. α is a "step size" parameter that potentially can be tuned. Notice, when $\alpha \to 0$ only very small steps in the negative gradient direction are taken. In [16] it was proven that for the least squares error $\alpha = 1$ is guaranteed to monotonically decrease the cost function.

The shifted non-negative matrix factorization model proposed in [17] is given by

$$x_{m,n} \approx \sum_d a_{m,d} s_{d,n-\tau_{m,d}},$$

where **X**,**A** and **S** are non-negative. While shifts correspond to simple multiplication of a complex phase, the non-negativity constraint is not transparent in the frequency domain. Thus, a method combining the apparent representation of non-negativity in the time domain with the efficient implementation of shifts in the frequency domain is desired. The least squares objective can be written as

$$C_{LS}(A,S) = \frac{1}{2N}\sum_f \|\widetilde{\mathbf{x}}_f - \widetilde{\mathbf{A}}^{(f)}\widetilde{\mathbf{s}}_f\|_F^2.$$

Thus, in the frequency domain the objective becomes separable over frequencies, however the non-negativity constraint resides in the time domain. The model can be estimated alternatingly solving for **A**, **S** and $\boldsymbol{\tau}$ as described in [17]. Here, the TFGM is applied for update of the variable **S**. The gradient of the least squares cost function in the frequency domain is [18]

$$\widetilde{\mathbf{G}}_f = \frac{\partial C_{LS}}{\partial \widetilde{\mathbf{H}}_f} = -\frac{1}{N}\widetilde{\mathbf{A}}^{(f)H}(\widetilde{\mathbf{x}}_f - \widetilde{\mathbf{A}}^{(f)}\widetilde{\mathbf{s}}_f).$$

By applying the inverse DFT of the gradient in the frequency domain the corresponding gradient in the time domain is obtained. Splitting the gradient in the frequency domain into what constitutes the positive and negative part of the corresponding gradient in the time-domain gives

$$\begin{aligned}\widetilde{\mathbf{G}}_f^+ &= \frac{1}{N}\widetilde{\mathbf{A}}^{(f)H}\widetilde{\mathbf{A}}^{(f)}\widetilde{\mathbf{s}}_f,\\ \widetilde{\mathbf{G}}_f^- &= \frac{1}{N}\widetilde{\mathbf{A}}^{(f)H}\widetilde{\mathbf{x}}_f.\end{aligned}$$

Consequently, by applying the inverse DFT to $\widetilde{\mathbf{G}}_f^+$ and $\widetilde{\mathbf{G}}_f^-$ the corresponding positive and negative parts of the gradient in the time-domain are found. As a result, **S** can be updated using multiplicative updates in the time domain, hence, enforcing non-negativity through the update

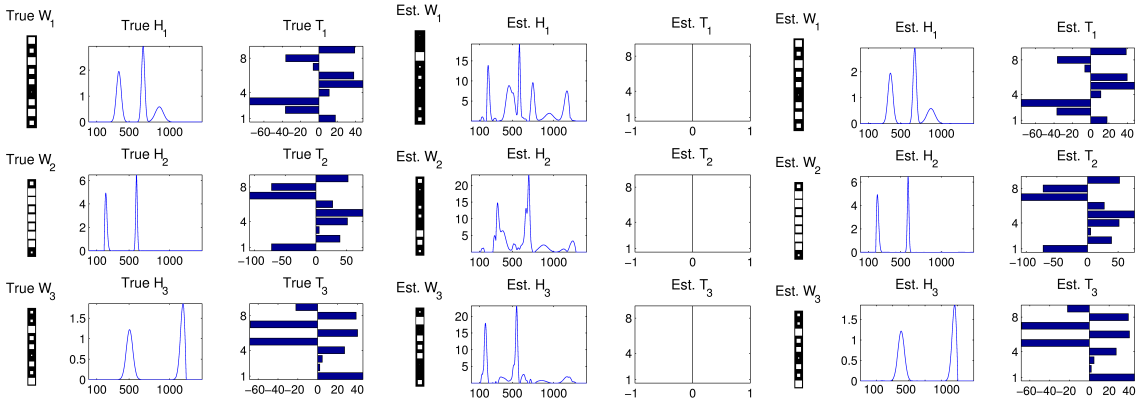
$$s_{d,n} = s_{d,n}\left(\frac{g_{d,n}^-}{g_{d,n}^+}\right)^{\alpha}.$$

In Figure 1 we demonstrate the usefulness of the ShifNMF over regular instantaneous NMF when shifts are present in the data.

3.2. Convolutive Sparse Coding

The Convolutive Sparse Coding model is given by

$$x_{m,n} \approx \sum_{d,\tau} a_{m,d}^{\tau} s_{d,n-\tau}.$$



(a) True components (b) Estimated components NMF (c) Estimated components ShiftNMF

Fig. 1. **Left panel:** The true factors forming the synthetic data ($\mathbf{X} \in \Re^{9\times 1400}$). To the left, the strength of the mixing $\mathbf{A}$ of each source is indicated in gray color scale. In the middle, the three sources are shown and to the right is given the time delays of each source to each channel. **Middle panel:** Results obtained by conventional instantaneous NMF for the synthetic data given in figure 3. Clearly, the model can not account for the shifts in the data hence the sources are incorrectly estimated. Notice, only 68 % of the variance of the data can be accounted for. **Right panel:** The estimated factors obtained by a ShiftNMF analysis. Clearly, the model with shifts has correctly recovered the components of the synthetic data hence accounts for all the variance in the data.

Where $\mathbf{S}$ is sparse. The model is separable in the frequency domain and can be optimized using the following objective of the form given in (4)

$$C = \frac{1}{2N}\sum_f \|\widetilde{\mathbf{x}}_f - \widetilde{\mathbf{A}}_f\widetilde{\mathbf{s}}_f\|_F^2 - \lambda\sum_n log(sp(\mathbf{s}_n))$$

Where the first term is the reconstruction error and second term the sparsity penalty imposed with strength λ given by the sparse prior distribution sp. We will presently consider the Laplace prior given by $sp(\mathbf{s}_n) = e^{-|\mathbf{s}_n|}$ forming a l_1-norm regularization penalty. As for non-negativity constraints, the sparsity in the time domain as well as regions where the filter $\mathbf{A}^\tau$ is zero is not transparent in a frequency domain representation. However, the convolutive model is efficiently estimated in a frequency domain representation. Thus, again the TFGM admits the benefits of the representations in the two domains. The gradient of the least squares error in the frequency domain is given by

$$\begin{aligned}\nabla^{LS}_{\widetilde{a}_{m,d,f}} &= -\frac{1}{N}(\widetilde{x}_{m,f} - \sum_d \widetilde{a}^*_{m,d,f}\widetilde{s}_{d,f})\widetilde{s}^*_{d,f}\\ \nabla^{LS}_{\widetilde{s}_{d,f}} &= -\frac{1}{N}\sum_m \widetilde{x}_{m,d,f}(\widetilde{x}_{m,f} - \sum_d \widetilde{a}^*_{m,d,f}\widetilde{s}_{d,f})\end{aligned}$$

Thus, the gradient in the time domain is given by

$$\begin{aligned}\nabla_{\mathbf{A}^\tau} &= \mathcal{F}^{-1}(\nabla^{LS}_{\widetilde{\mathbf{A}}})_\tau\\ \nabla_{\mathbf{S}} &= \mathcal{F}^{-1}(\nabla^{LS}_{\widetilde{\mathbf{S}}}) + \lambda\,\mathrm{sign}(\mathbf{S})\end{aligned}$$

Hence, by computing the gradient in the time domain it becomes transparent how $\mathbf{A}^\tau$ can be estimated such that only active regions of the filter $\mathbf{A}^\tau$ are updated. Furthermore, the update in the time domain of $\mathbf{S}$ enables the combination of sparseness constraint in the time domain with efficient representation in the frequency domain. $\mathbf{A}^\tau$ and $\mathbf{S}$ are updated using linesearch, i.e., by $\mathbf{A}^\tau \leftarrow \mathbf{A}^\tau - \mu_A\nabla_{\mathbf{A}^\tau}$ and $\mathbf{S} \leftarrow \mathbf{S} - \mu_S\nabla_{\mathbf{S}}$.

In Figure 2 we demonstrate the proposed algorithm on an EEG-data set described in [19] based on a visual paradigm. We removed the three frontal electrodes EOG1, EOG2 and FPz heavily confounded by eye artifacts prior to the analysis.

3.3. Sparse and Smooth Matrix Factorization

Smoothness constraints in the time domain corresponds to reduced high frequency content. Hence smoothness can be imposed by penalizing high frequency regions of the components, i.e., by considering the following objective in the form given in (4)

$$C = \sum_n \frac{1}{2}\|\mathbf{x}_n - \mathbf{A}\mathbf{s}_n\|_F^2 + \frac{\lambda_2}{2N}\sum_f w_f\|\widetilde{\mathbf{s}}_\mathbf{f}\|_F^2.$$

From the objective above it can be seen that smoothness does not improve the identifiability of the model since multiplying the sources $\mathbf{S}$ by the orthogonal matrix $\mathbf{Q}$ result in a representation that is equally smooth, i.e. $\|\widetilde{\mathbf{s}}_\mathbf{f}\|_F = \|\mathbf{Q}\widetilde{\mathbf{s}}_\mathbf{f}\|_F$. Thus, additional constraints

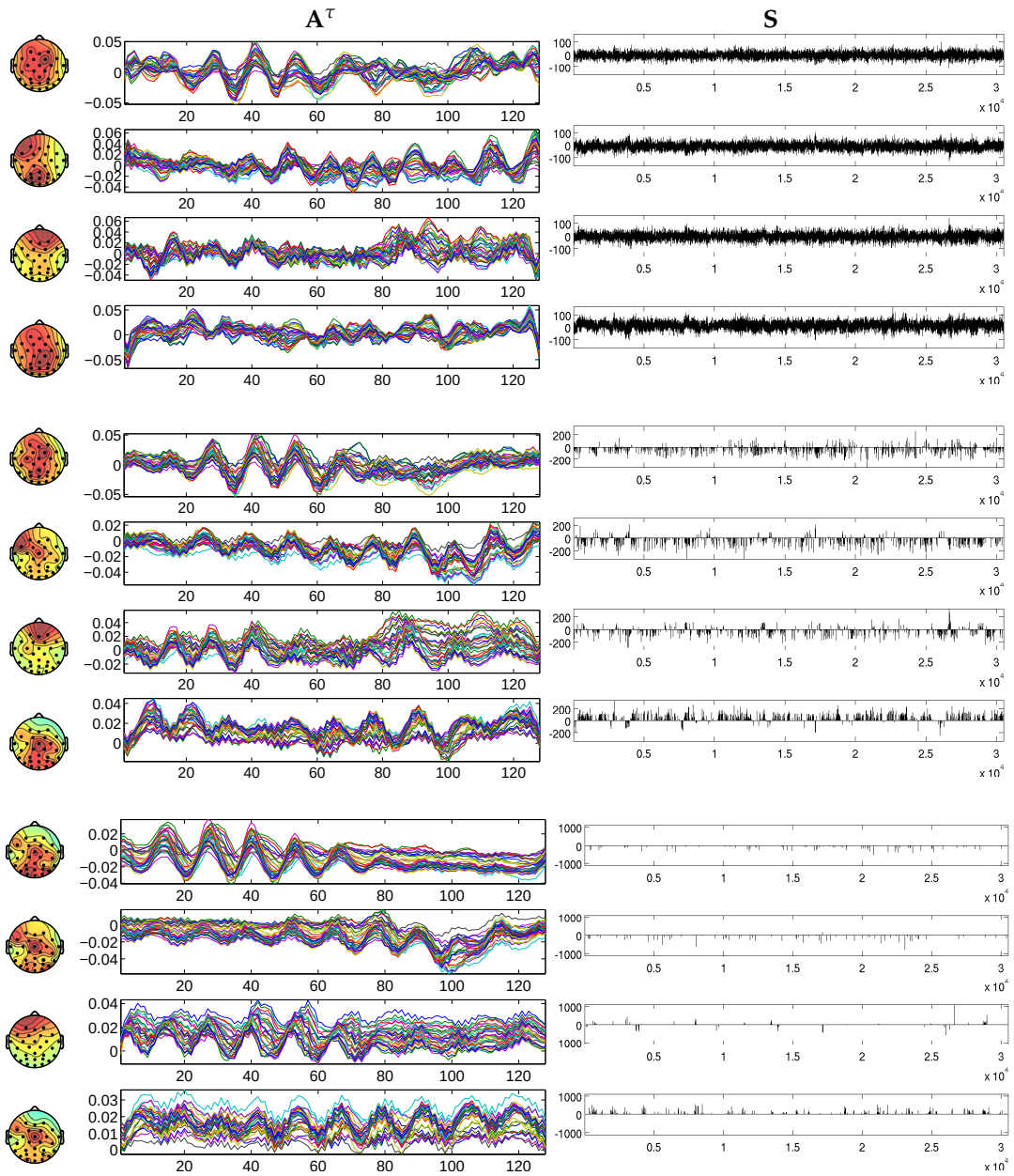


Fig. 2. Convolutive Sparse Coding analysis of EEG data obtained from a visual paradigm. The size of the data is $\mathbf{X} \in \Re^{29\times 30504}$ while $\mathbf{A}^\tau \in \Re^{29\times 4}$ and $\tau \in [1, 2, \ldots, 128]$. **Top panel:** Analysis for $\lambda = 0$, clearly $\mathbf{S}$ given to the right is not sparse thus the EEG activity is modeled both in the convolutive filter $\mathbf{A}^\tau$ and in the sources $\mathbf{S}$. The scalp maps to the left gives the power of the filter coefficients for the electrodes of each component. The explained variation is 91% **Middle panel:** Analysis based on $\lambda = 200$, clearly $\mathbf{S}$ has become sparse while the temporal structure of the EEG data mainly is coded in the filter $\mathbf{A}^\tau$. The explained variation is 66%. **Bottom panel:** When increasing the sparsity strength ($\lambda = 700$) $\mathbf{S}$ becomes even more sparse. The explained variation is 35%. The activity captured by the models are mainly the powerful alpha activity residing in a frequency band around 8-12 Hz.

are necessary in order to obtain an unambiguous representation. We will here improve identifiability of the model by imposing sparseness on $\mathbf{S}$. Again, sparsity is not transparent in a frequency representation. However, the sparsity and smoothness constraints can again be combined using the proposed TFGM. Consider the following sparse and smooth matrix factorization

$$C = \frac{1}{2}\sum_n (\|\mathbf{x}_n - \mathbf{A}\mathbf{s}_n\|_F^2 + \lambda_1\|\mathbf{s}_n\|_1) + \frac{\lambda_2}{2N}\sum_f w_f\|\widetilde{\mathbf{s}}_\mathbf{f}\|_F^2$$

Where w_f weights frequencies according to the smoothness desired. Clearly, the objective has the form given in (4). Thus, the gradient of the above objective is given by

$$\begin{aligned}\nabla_\mathbf{A} &= -(\mathbf{A}\mathbf{S} - \mathbf{X})\mathbf{S}^T \\ \nabla_\mathbf{S} &= -\mathbf{A}^T(\mathbf{A}\mathbf{S} - \mathbf{X}) + \lambda_1 sign(\mathbf{S}) + \lambda_2\mathcal{F}^{-1}(\widetilde{\mathbf{S}}').\end{aligned}$$

where, $\mathbf{s}'_f = w_f\mathbf{s}_f$. Again $\mathbf{A}$ and $\mathbf{S}$ are updated using line-search, i.e. by $\mathbf{A} \leftarrow \mathbf{A} - \mu_A\nabla_\mathbf{A}$ and $\mathbf{A} \leftarrow \mathbf{S} - \mu_S\nabla_\mathbf{S}$.

In Figure 3 we demonstrate the algorithms ability to impose smoothness on simulated data while at the same time obtain better identifiability through sparseness.

4. DISCUSSION AND CONCLUSION

We have proposed the Time Frequency Gradient Method that takes advantage of mixing representations which allows us to handle convolutions and shifts efficiently while imposing time domain constraints. The framework can be used for objective functions that are separable in time or frequency parts such that variables can be arbitrarily updated in the time or frequency domain. We demonstrated the viability of this simple approach for models that involve shifts, convolutions, or smoothness, but we expect the framework to be useful for a wide range of models where the frequency representation facilitates efficient computation.

5. REFERENCES

[1] Charles Spearman, ""general intelligence," objectively determined and measured," *American Journal of Psychology*, vol. 15, pp. 201–293, 1904.

[2] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrica*, vol. 23, pp. 187–200, 1958.

[3] Pierre Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[4] Anthony J. Bell and Terrence J. Sejnowski, "An information maximization approach to blind source separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[5] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[6] R.A. Harshman, S. Hong, and M.E. Lundy, "Shifted factor analysis part i: Models and properties," *Journal of Chemometrics*, vol. 17, pp. 363–378, 2003.

[7] K Torkkola, "Blind separation of delayed sources based on informationmaximization," *Acoustics, Speech, and Signal Processing. ICASSP-96*, vol. 6, pp. 3509–3512, 1996.

[8] Arie Yeredor, "Time-delay estimation in mixtures," *ICASSP*, vol. 5, pp. 237–240, 2003.

[9] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution: the dynamic component analysis algorithm," *Neural Computation*, vol. 10, no. 6, pp. 1373–1424, 1998.

[10] L. Parra, C. Spence, and B.D. Vries, "Convolutive blind source separation based on multiple decorrelation," *IEEE Workshop on Neural Networks and Signal Processing*, pp. 23–32, 1998.

[11] J. Anemuller, Terrence J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, 2003.

[12] T. Blumensath and M. Davies, "On shift-invariant sparse coding," *International Conference on Independent Component Analysis and Blind Source Separation*, vol. 26, pp. 1205–1212, 2004.

[13] M. S. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse shift-invariant representations," *Adv. Neural Inform. Process. Systems (NIPS'99)*, vol. 11, pp. 730–736, 1999.

[14] Michael Pedersen Syskind, Jan Larsen, Ulrik Kjems, and Lucas C. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook on Speech Processing and Speech Communication*, 2007.

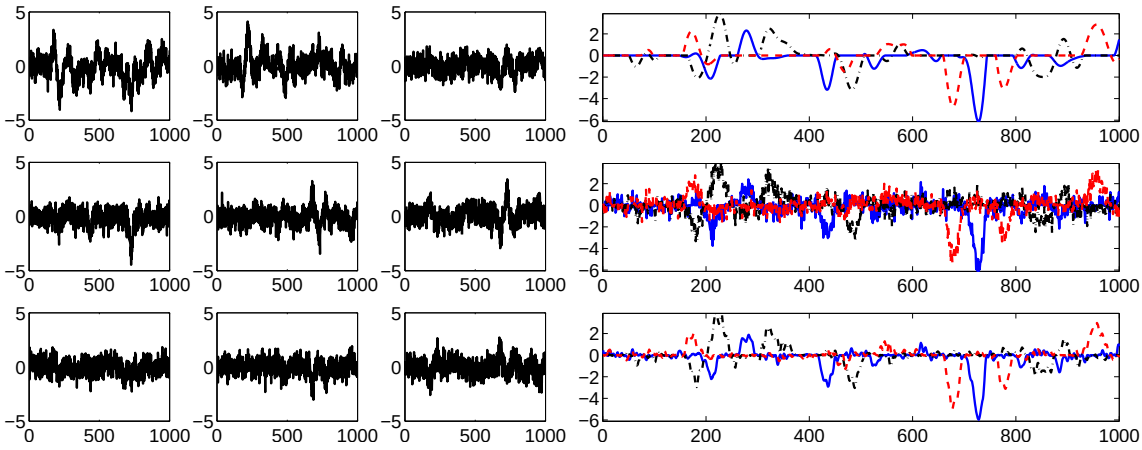


Fig. 3. **Left Panel:** The nine realizations of the simulated input data ($X \in \Re^{9\times 1000}$) based on a mixture of the three sources given in the top right panel with additive Normal distributed noise. Notice that the signal is heavily contaminated with a signal to noise ratio of 0 db. **Right Panel:** The top figure show the three true underlying sources used to simulate data, all three components are both sparse and smooth. The middle figure shows the sources identified by the proposed smooth and sparse matrix factorization algorithm imposing ($\lambda_1 = 0.25$, $\lambda_2 = 0$, i.e., without any smoothness regularization). The bottom figure shows how introducing smoothness regularization suppresses noise in the solution ($\lambda_1 = 0.25$, $\lambda_2 = 20$ for frequencies larger than 0.14 samples^{-1}). While smoothness suppresses noise in the sources it will also introduce a bias in this particular case reducing the explained variation from 64% to 53%.

[15] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–91, 1999.

[16] Daniel D Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *NIPS*, 2000, pp. 556–562.

[17] M. Mørup, K. H. Madsen, and L. K. Hansen, "Shifted non-negative matrix factorization," 2007, pp. 139–144.

[18] Kaare B. Petersen and Michael S. Pedersen, "The matrix cookbook," www.matrixcookbook.com, 2007.

[19] S. Makeig, M. Westerfield, T.P. Jung, J. Covington, J. Townsend, T.J. Sejnowski, and E. Courchesne, "Functionally independent components of the late positive event-related potential during visual spatial attention," *J. Neurosci.*, vol. 19, pp. 2665–2680, Apr 1999.